Online Data Supplement

Title: Genome-wide association study in patients with pulmonary *Mycobacterium avium* complex disease

Authors

Ho Namkoong [1)2)3)], Yosuke Omae [4)5)], Takanori Asakura [1)6)], Makoto Ishii [1)], Shoji Suzuki [1)], Kozo Morimoto [7)] , Yosuke Kawai [4)5)], Katsura Emoto [8)], Andrew J. Oler [9)], Eva P. Szymanski [2)], Mitsunori Yoshida [6)], Shuichi Matsuda [7)], Kazuma Yagi [1)], Isano Hase [10)], Tomoyasu Nishimura [11)], Yuka Sasaki [7)], Takahiro Asami [1)], Tetsuya Shiomi [12)], Hiroaki Matsubara [13)], Hisato Shimada [14)], Junko Hamamoto [1)], Byung Woo Jhun [15)], Su-Young Kim [15)], Hee Jae Huh [16)], Hong-Hee Won [17)], Manabu Ato [6)], Kenjiro Kosaki [18)], Tomoko Betsuyaku [1)], Koichi Fukunaga [1)], Atsuyuki Kurashima [7)], Hervé Tettelin [19)20)], Hideki Yanai [21)], Surakameth Mahasirimongkol [22)], Kenneth N. Olivier [23)], Yoshihiko Hoshino [6)], Won-Jung Koh [15)], Steven M. Holland [2)], Katsushi Tokunaga [4)5)], Naoki Hasegawa [24)]

Namkoong H and Omae Y are co-first authors.

1) Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
2) Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Maryland, USA
3) JSPS Overseas Research Fellow, Japan Society for the Promotion of Science, Tokyo, Japan.
4) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

5) Genome Medical Science Project (Toyama), National Center for Global Health and Medicine, Tokyo, Japan
6) Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Tokyo, Japan
7) Respiratory Disease Center, Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan
8) Department of Pathology, Keio University School of Medicine, Tokyo, Japan
9) Bioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Maryland, USA
10) Department of Respiratory Medicine, National Hospital Organization Utsunomiya Hospital, Tochigi, Japan
11) Keio University Health Center, Tokyo, Japan
12) Department of Pulmonary Medicine, Keiyu Hospital, Kanagawa, Japan
13) Department of Pulmonary Medicine, Fussa Hospital, Tokyo, Japan
14) Department of Pulmonary Medicine, Kawasaki Municipal Ida Hospital, Kanagawa, Japan
15) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea
16) Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea
17) Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, South Korea
18) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan
19) Department of Microbiology and Immunology, School of Medicine, University of Maryland, Maryland, USA
20) Institute for Genome Sciences, School of Medicine, University of Maryland, Maryland, USA

21) Department of Clinical Laboratory, Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan

22) Medical Genetics Center, Medical Life Sciences Institute, Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand

23) Pulmonary Branch, National Heart, Lung, and Blood Institute, National Institutes of Health (NIH), Maryland, U.S.A.

24) Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan

* Nontuberculous Mycobacteriosis Japan Research Consortium (NTM-JRC)

Correspondence:

Ho Namkoong M.D., Ph.D., MPH

Laboratory of Clinical Immunology and Microbiology, LCIM

National Institute of Allergy and Infectious Diseases, National Institutes of Health

9000 Rockville Pike, Building 10, Room 11S261

Bethesda, MD 20814, USA

Phone: 301-761-7501

Email: ho.namkoong@nih.gov hounamugun@gmail.com

Naoki Hasegawa M.D. Ph.D.

Department of Infectious Diseases

Keio University School of Medicine

35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

Phone: +81-3- 5363-3710(63875)

E-mail: n-hasegawa@z8.keio.jp

Katsushi Tokunaga, Ph.D.

Genome Medical Science Project (Toyama)

National Center for Global Health and Medicine

Toyama 1-21-1, Shinjuku-ku, Tokyo 162-8655, Japan

E-mail: katokunaga@ri.ncgm.go.jp

Supplemental Methods

GWAS genotyping and data quality control

Genome-wide SNP genotyping was conducted in 549 patients with pulmonary NTM and 419 healthy controls using the Affymetrix Axiom Japonica array (659,636 markers for 659,503 SNPs; Santa Clara, CA, USA). One sample was excluded because it showed a dish quality control value of <0.82, which is a recommended quality metric for the Axiom array. The genotypes of the remaining samples were called using Genotyping Console v4.2 software and 10 cases and 2 controls with an overall call rate of <97% were removed. Identity by descent testing was used to analyse 538 cases and 417 controls to determine cryptic relatedness, and 9 duplicate or related samples were removed. All remaining sample pairs showed relatedness (PI_HAT) values of <0.125, which is the expected value for third-degree relatives; this indicated that the remaining samples were from unrelated individuals. Principal component analysis using the public HapMap Phase III data as controls revealed that four cases belonged to Han Chinese in Beijing population; these cases were removed from subsequent analysis (Supplementary Figure S1). In total, 525 cases and 417 controls passed GWAS quality control; 14 suspicious NTM cases and 36 non-MAC cases were excluded prior to analysing the associations between genetic variants and MAC susceptibility.

The following three quality control thresholds were employed for SNP genotyping: SNP call rate ≥95%; MAF ≥1%; and Hardy-Weinberg equilibrium p-value ≥0.001 in healthy controls. These thresholds excluded 8,756, 9,656, and 4,266 of the 645,712 SNPs analysed on autosomal

chromosomes, respectively. All cluster plots for SNPs with p-values <1E-04 (chi-squared test of the allele frequency model) were evaluated by visual inspection, and 66 SNPs with ambiguous genotype calls were excluded. In total, 622,723 autosomal SNPs passed quality control and were used for whole genome genotype imputation.

Whole genome imputation and SNP selection for replication study

Genotype imputation was conducted with IMPUTE4 software using a phased reference panel of 2,049 Japanese individuals from a prospective, general population cohort study performed by the Tohoku Medical Megabank Organization (ToMMo). Pre-phasing was conducted with EAGLE v2.3.242 with default settings. For quality control of the imputed genotypes, the imputation probability threshold of 0.9 recommended by the developer was applied, and SNPs with > 5% un-imputed genotype data, an MAF < 1%, or a Hardy-Weinberg equilibrium p-value < 0.001 in controls were excluded from the subsequent association analysis. Regional association plots were constructed by LocusZoom. Candidate SNPs with a p-value threshold of 3E-06 in genome-wide association study (GWAS) were extracted and genotyped SNPs in LD ($r2 > 0.7$) with the candidate SNPs were selected for replication study (Supplementary Table S1).

Conditional analysis

Conditional analysis was performed using PLINK software to identify independent signals at the chromosome 16p21 locus in the GWAS discovery dataset or in the combined dataset.

Conditional associations of all variants located within 400 kbp of the candidate SNP were analysed in the discovery dataset.

eQTL analysis

Data relating to the correlations between candidate SNP genotypes and surrounding gene expression were available via the Genotype-Tissue Expression (GTEx) project portal database (http://gtexportal.org/home/; Release v8). Significance associations were based on a Q-value threshold provided in GTEx database.

Fine-mapping analysis

We performed fine-mapping of candidate loci including variants with suggestive evidence of association in our discovery dataset. We employed FINEMAP v1.3.1 software and targeted variants within 1Mbp region of the most significant SNP. Beta values and their standard errors were calculated based on the odds ratio (OR) and their 95% lower confidence interval, and linkage disequilibrium information among variants was generated using matrix option of PLINK software. The shotgun stochastic search was applied to detect 95% credible variant set to explain the association in our discovery dataset.

Colocalisation analysis

We performed Bayesian colocalisation analyses between identified signal in our GWAS and eQTL associations in GTEx dataset. GTEx provides 49 human tissue eQTL data from 838

individuals in its v8 release. We employed the COLOC package in R that tests for five hypotheses in the candidate region: H0, neither GWAS or eQTL has a genetic association; H1, only GWAS has a genetic association; H2, only eQTL has a genetics association; H3, both GWAS and eQTL are associated, but with different causal variants; and H4, both GWAS and eQTL are associated and share a causal variant. In practice, a high posterior probability of H4 (PP4>80%) is considered as evidence for colocalisation. For our candidate GWAS signal at chromosome 16p21, we defined a candidate region including SNPs with p<1E-04 (chr16:23,742,273- 23,814,204 in hg38). We next extracted all tissue-transcript pairs that can capture our candidate region within their cis-eQTL window (a plus and minus 1Mb around the transcription start site of the transcript as defined by GTEx) using egene file provided by GTEx, and searched significant variant-transcript eQTL pairs located in our candidate region. If any overlapping significant eQTLs were detected, we tested for the colocalisation using all variant-transcript associations data for the candidate tissue-transcript pair downloaded from google cloud platform provided by GTEx. The effect alleles for GWAS and eQTL were harmonized based on the hg38 positions and alternative/effect allele information in lookup table provided by GTEx. Then their posterior probabilities of colocalisation were calculated using the "coloc.abf" function with default parameters.

Immunostaining of resected lung tissue

Lung tissues were obtained from patients who underwent pulmonary resection for pulmonary MAC disease and patients with lung cancers. Pulmonary tissues were fixed in formalin (10%).

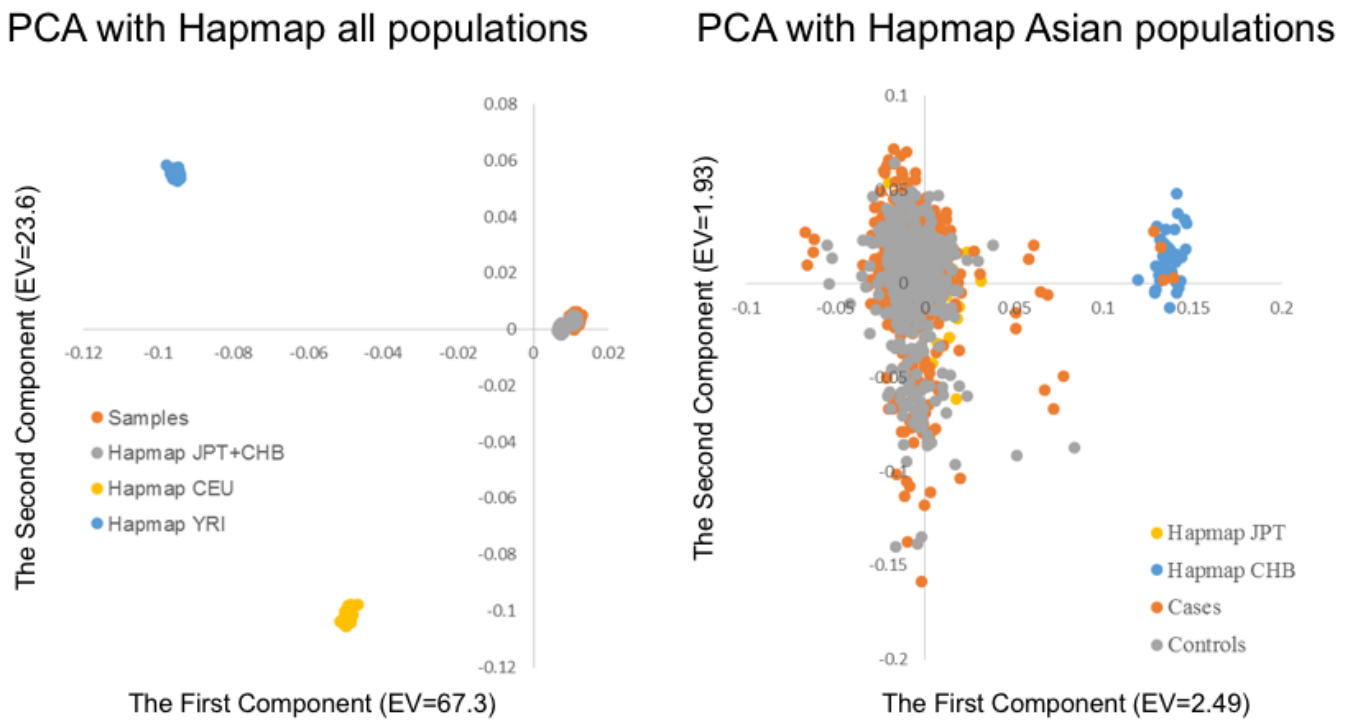
Sliced tissues were embedded in paraffin. Formalin-fixed, paraffin-embedded sections were immunostained with an anti-CHP2 monoclonal antibody (1: 100; Signalway Antibody, Maryland, USA). Immunostaining was performed with the Leica Bond-Max automation and Leica Refine detection kit (Leica Biosystems, Bannockburn, IL). Since smooth muscle and bronchiolar respiratory epithelia are stained weakly among patients both with or without pulmonary MAC, we set those as inner control (Supplementary Figure S6). We defined negative for weaker than that of inner control, equivocal for equal as that of inner control level and positive for stronger than that of inner control.

Statistical analysis

For the GWAS and imputation analysis, a chi-squared test was applied to a two-by-two contingency table in an allele frequency model. A quantile-quantile plot of the distribution of test statistics showed that its genomic inflation factor was 1.033. The genome-wide significance threshold was set at 5E-08, and SNPs with a p-value < 1E-05 and an FDR < 0.50 were considered to be suggestive of the genetic association in this study. For verification cohorts of populations other Japanese and other diseases, SNPs with a p-value < 5E-02 were considered to be significant. Demographic data were compared using one-way ANOVA or a chi-squared test. All analyses were performed using GraphPad Prism7. Two-sided p-values of < 0.05 were considered statistically significant.

Ethics Statement

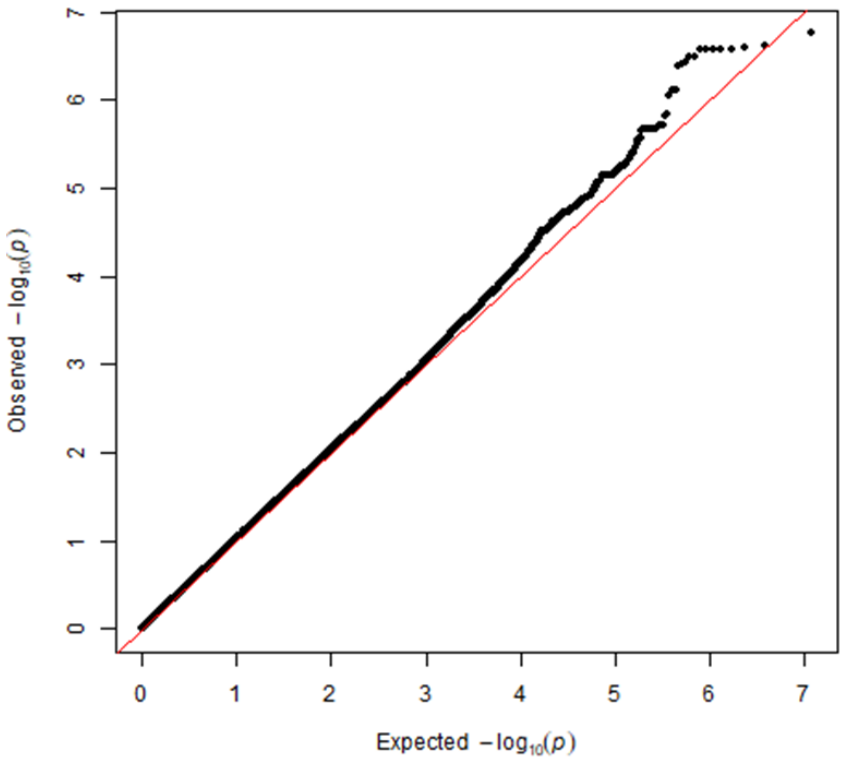
This study was reviewed and approved by the research ethics committees of the Keio University School of Medicine (2012-0336), the Graduate School of Medicine, the University of Tokyo (G3579), and other collaborating institutions (UMIN000021692). Regarding the European cohort, the subjects were enrolled in studies under the National Institute of Allergy and Infectious Diseases Institutional Review Board-approved protocol (93-I-0119), the National Heart, Lung and Blood Institutional Review Board-approved protocol (09-H-0172). Regarding the Korean cohort, the subjects were enrolled in studies under the Institutional Review Board of Samsung Medical Center (IRB No. 2008-09-016). All adult subjects provided written informed consent and all the experiments were performed in accordance with the relevant guidelines and regulations.



Supplementary Figure S1. Principal component analysis.

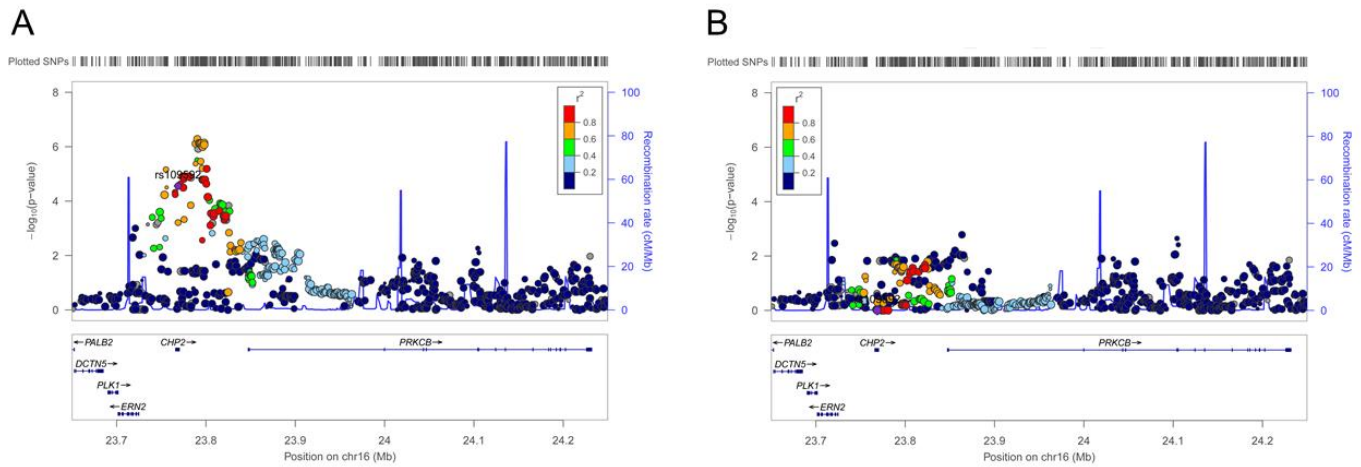
(A) Analysis of HapMap samples from all populations. The study samples clustered with HapMap Asian samples. (B) Analysis with HapMap Asian population samples. Four MAC samples clustered with the HapMap CHB samples and were therefore removed. The GWAS included 892 samples (475 MAC cases and 417 controls) that clustered with the HapMap JPT samples.

PCA: principal component analysis, JPT: Japanese in Tokyo, CHB: Han Chinese in Beijing, CEU: Utah residents with Northern and Western European ancestry from the CEPH collection, YRI: Yoruba in Ibadan



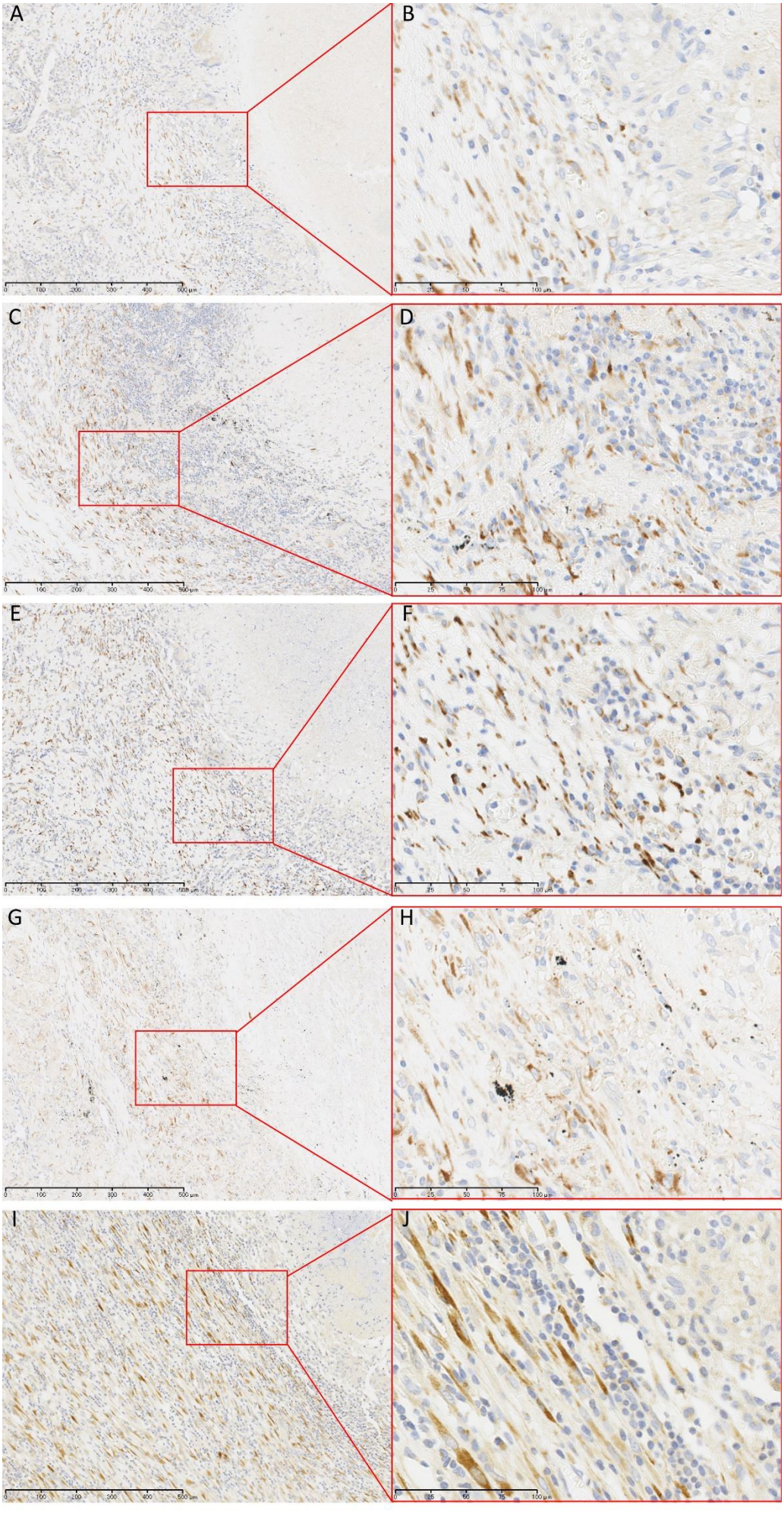
Supplementary Figure S2. Quantile-Quantile plot.

The genomic inflation factor lambda was 1.033 for the allelic model.



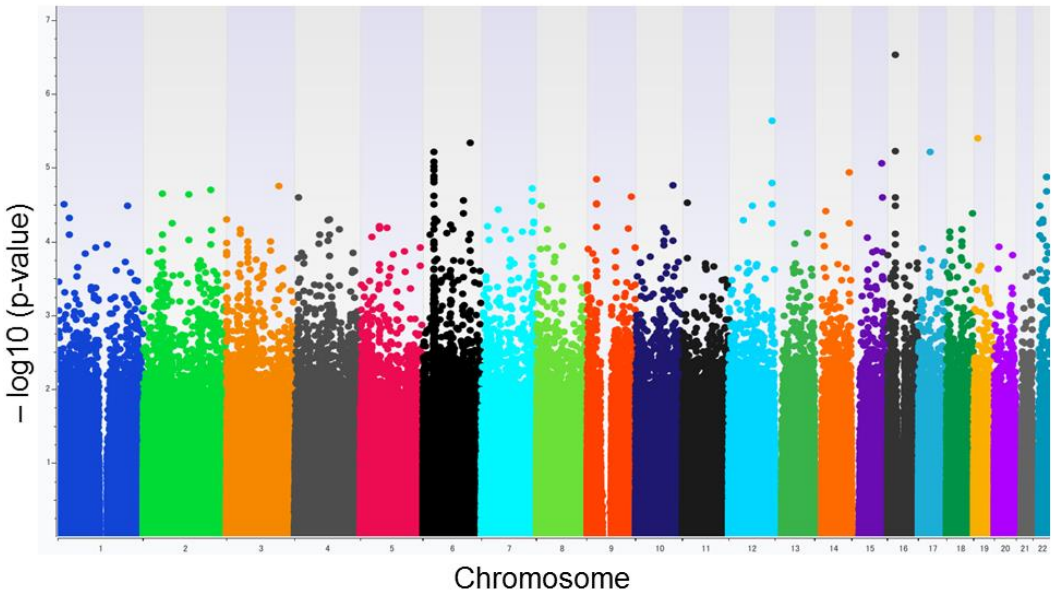
Supplementary Figure S3. Conditional analysis of chromosome 16p21 using rs109592.

The associations using the logistic regression model are shown (A) before and (B) after conditioning with rs109592. The lowest p-values before and after this conditional analysis were 5.21E-07 (for rs194788) and 1.64E-03 (for rs148351442), respectively.


A
B
C
D
E
F
G
H
I
J

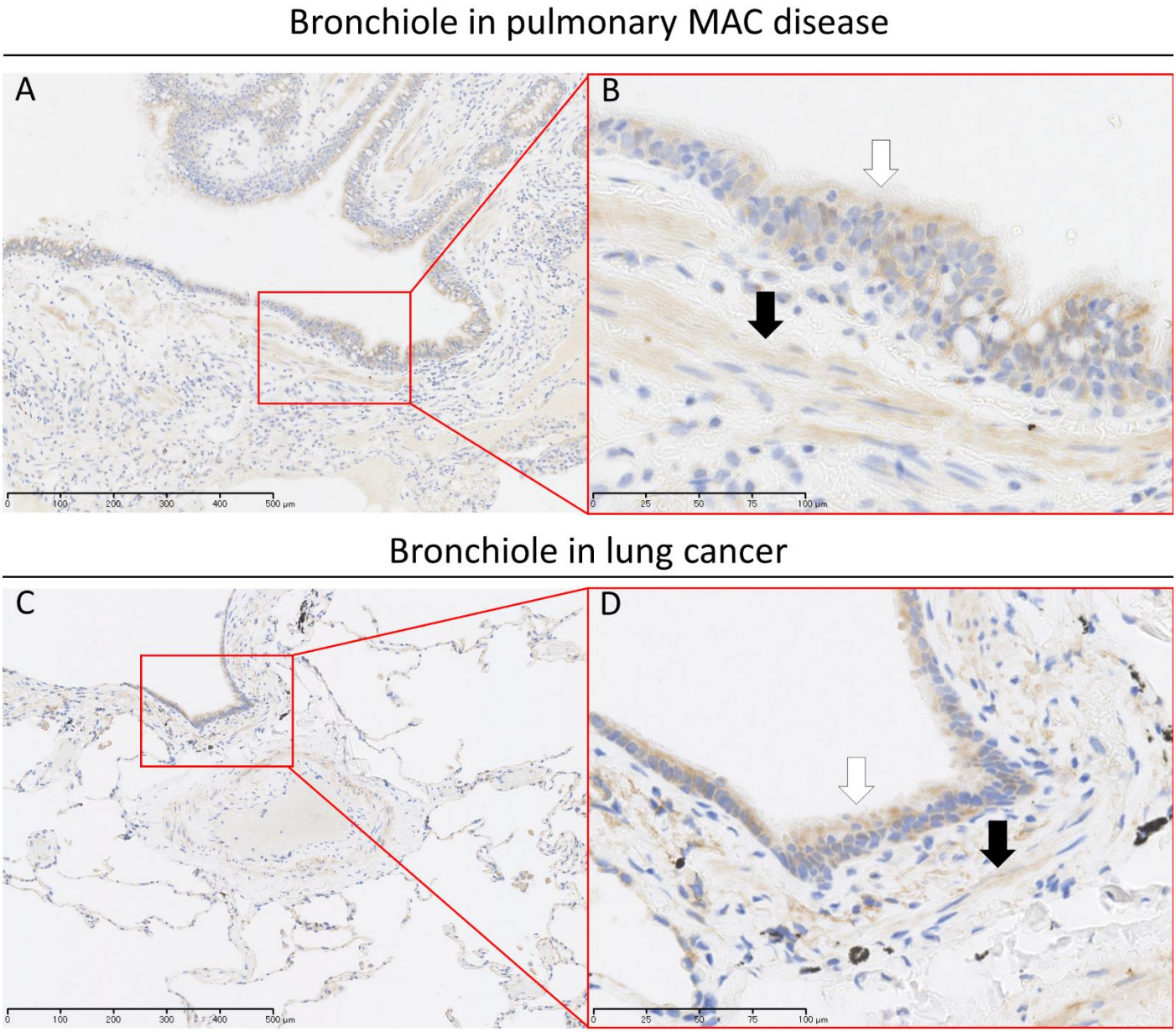
Supplementary Figure S4. Immunostaining of resected lung tissue from pulmonary MAC patients.

The magnification is (A)(C)(E)(G)(I) ×100 and (B)(D)(F)(H)(J) ×400. CHP2 was positive with fibroblasts surrounding granuloma in all cases(A-J).



Supplementary Figure S5. Manhattan plot of GWAS data.

Data are shown for 622,723 autosomal SNPs in 511 Japanese pulmonary NTM cases and 417 Japanese healthy controls. The y axis shows –log10 p-values of 622,723 autosomal SNPs, and x axis shows their chromosomal positions.



Supplementary Figure S6. Immunostaining of resected lung tissue.

The magnification is (A)(C) ×100 and (B)(D) ×400. Smooth muscle (black arrow) and bronchiolar respiratory epithelia (white arrow) are stained weakly among patients both with (A)(B) or without (C)(D) pulmonary MAC.

Supplementary Table S1. Candidate SNPs from GWAS discovery stage

RSID	Chr	Position (hg19)	Minor Allele	Major Allele	Discovery stage (475 MAC vs 417 Controls)						Regulome DB Score (v 1.1)
					MAF (Cases)	MAF (Controls)	OR (95% CI)	P	Info score	LD with Tag SNP (r2)	
rs11646605	**16**	**23,793,439**	**T**	**G**	**0.164**	**0.264**	**0.55 (0.44-0.69)**	**2.67E-07**	**1.000**	**Tag / genotyped**	**6**
rs194788	16	23,790,338	T	A	0.163	0.264	0.54 (0.43-0.68)	1.75E-07	0.997	1.000	6
rs185452	16	23,791,043	T	A	0.152	0.251	0.53 (0.42-0.68)	2.38E-07	0.977	0.976	7
rs194790	16	23,792,157	G	A	0.164	0.263	0.55 (0.43-0.69)	2.57E-07	0.998	1.000	5
rs2520013	16	23,793,850	G	A	0.164	0.264	0.55 (0.44-0.69)	2.67E-07	1.000	1.000	6
rs12051520	16	23,795,244	G	A	0.164	0.264	0.55 (0.44-0.69)	2.67E-07	1.000	1.000	7
rs2575385	16	23,797,813	G	A	0.165	0.265	0.55 (0.44-0.69)	2.71E-07	0.998	0.994	6
rs2520020	16	23,797,876	G	A	0.165	0.265	0.55 (0.44-0.69)	2.71E-07	0.999	0.994	7
rs2520021	16	23,798,046	G	A	0.165	0.264	0.55 (0.44-0.69)	3.17E-07	0.998	0.994	5
rs12449160	16	23,788,291	G	A	0.162	0.261	0.55 (0.43-0.69)	3.22E-07	0.995	1.000	7
rs2520016	16	23,795,923	C	A	0.165	0.264	0.55 (0.44-0.70)	3.66E-07	0.998	0.990	7
rs9923292	16	23,796,375	G	A	0.163	0.262	0.55 (0.44-0.69)	3.83E-07	0.992	0.994	4
rs142688521	16	23,791,217	GATA	G	0.164	0.262	0.55 (0.44-0.70)	4.14E-07	0.994	1.000	6

rs11647385	16	23,789,907	G	T	0.153	0.246	0.56 (0.44-0.71)	1.47E-06	0.974	0.989	6
rs2520014	16	23,794,098	A	G	0.162	0.255	0.57 (0.45-0.71)	1.48E-06	0.990	0.993	6
rs12449169	16	23,788,358	G	A	0.156	0.246	0.56 (0.44-0.72)	2.22E-06	0.980	0.993	6
rs183172	16	23,800,830	C	T	0.125	0.208	0.55 (0.42-0.70)	2.70E-06	0.996	0.731	7
rs1858874	16	23,795,844	G	A	0.164	0.254	0.57 (0.45-0.73)	2.86E-06	0.989	0.990	6
rs109592	**16**	**23,768,710**	**T**	**C**	**0.127**	**0.203**	**0.57 (0.44-0.74)**	**1.39E-05**	**0.992**	**0.741**	**1f**
rs7200798	**16**	**23,853,860**	**G**	**A**	**0.298**	**0.365**	**0.74 (0.60-0.90)**	**1.88E-03**	**0.992**	**0.422**	**1f**
rs7404095	16	23,864,590	T	C	0.302	0.372	0.73 (0.60-0.89)	2.75E-03	1.000	r2=1.000 with rs7200798	1d

rs12140922	**1**	**35,114,647**	**A**	**C**	**0.046**	**0.094**	**0.47 (0.32-0.69)**	**8.09E-05**	**1.000**	**Tag / genotyped**	**6**
rs12125323	1	35,099,742	G	A	0.037	0.091	0.38 (0.25-0.58)	2.71E-06	0.988	0.944	7

rs9668455	**12**	**126,702,012**	**C**	**T**	**0.168**	**0.260**	**0.58 (0.46-0.72)**	**2.17E-06**	**1.000**	**Tag / genotyped**	**7**
rs10847114	12	126,690,878	G	A	0.171	0.268	0.56 (0.45-0.71)	7.71E-07	0.984	0.983	7
rs10773246	12	126,691,312	G	A	0.171	0.268	0.56 (0.45-0.71)	7.71E-07	0.984	0.983	5
rs10773245	12	126,691,105	A	C	0.171	0.268	0.56 (0.45-0.71)	8.68E-07	0.982	0.980	5

rs7972004	12	126,695,369	G	T	0.168	0.261	0.57 (0.46-0.72)	1.89E-06	0.994	1.000	7
rs7975523	12	126,695,933	G	A	0.168	0.260	0.57 (0.45-0.72)	1.92E-06	0.992	1.000	6
rs7975685	12	126,701,346	C	T	0.169	0.262	0.57 (0.46-0.72)	2.10E-06	0.992	1.000	6
rs1029770	12	126,693,498	C	T	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.994	1.000	6
rs67196021	12	126,693,612	ACTTTC TT	A	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.994	1.000	6
rs2215302	12	126,694,217	C	T	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.994	1.000	6
rs2215303	12	126,694,232	A	G	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.994	1.000	6
rs10773247	12	126,694,826	C	T	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.995	1.000	5
rs10744243	12	126,694,907	A	G	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.995	1.000	5
rs10773248	12	126,696,886	A	G	0.168	0.260	0.58 (0.46-0.72)	2.17E-06	0.995	1.000	7
rs12894463	**14**	**98,999,346**	**C**	**A**	**0.265**	**0.180**	**1.65 (1.31-2.07)**	**1.65E-05**	**1.000**	**Tag / genotyped**	**7**
rs71416694	14	99,009,592	C	G	0.293	0.194	1.71 (1.37-2.14)	2.05E-06	0.988	0.883	3a

SNP: single nucleotide polymorphism, MAC: *Mycobacterium avium* complex, MAF: minor allele frequency, OR: odds ratio, CI: confidence interval, LD: linkage disequilibrium

Supplementary Table S2. SNPs previously reported to be associated with mycobacterial infections

Disease	Chr	Related Gene	RSID	Position (hg19)	Data	Minor allele	Major allele	MAF (Cases)	MAF (Controls)	OR (95%CI)	P
	6	HLA-DR-DQ	rs9271100	32,576,478	Genotyped	T	C	0.271	0.235	1.21 (0.97-1.50)	0.086
	6	HLA-DR-DQ	rs75324027	32,573,629	Imputed	G	A	0.428	0.404	1.10 (0.91-1.33)	0.28
	13	C13orf31	rs9567307	44,471,877	Imputed				QC failed		
	13	C13orf31	rs3764147	44,457,925	Genotyped	G	A	0.337	0.330	1.03 (0.85-1.26)	0.75
	16	NOD2	rs2270368	50,714,335	Imputed				QC failed		
	16	NOD2	rs9302752	50,719,103	Genotyped	C	T	0.273	0.237	1.20 (0.97-1.49)	0.089
	16	NOD2	rs16948876	50,854,418	Genotyped	A	G	0.043	0.038	1.13 (0.71-1.81)	0.61
	9	TNFSF15	rs4366152	117,564,875	Imputed	T	C	0.399	0.369	1.14 (0.94-1.38)	0.18
Leprosy	9	TNFSF15	rs6478108	117,558,703	Genotyped	C	T	0.401	0.373	1.13 (0.93-1.36)	0.22
	8	RIPK2	rs383592	90,804,328	Imputed	T	C	0.329	0.353	0.90 (0.74-1.09)	0.30
	8	RIPK2	rs160451	90,662,941	Imputed	T	C	0.190	0.176	1.10 (0.86-1.40)	0.51
	8	RIPK2	rs42490	90,778,513	Genotyped	A	G	0.333	0.356	0.90 (0.74-1.10)	0.30
	8	RIPK2	rs40457	90,823,687	Genotyped	G	A	0.215	0.227	0.93 (0.75-1.17)	0.55
	1	IL23R	rs2024825	67,597,977	Imputed				QC failed		
	1	IL23R	rs3762318	67,597,119	Genotyped	G	A	0.049	0.048	1.03 (0.67-1.59)	0.88
	2	IL18R1	rs76886731	102,945,416	Imputed	A	T	0.446	0.430	1.07 (0.88-1.29)	0.50
	6	RAB32	rs13220141	146,897,633	Imputed	T	A	0.167	0.159	1.06	0.74

										(0.82-1.36)	
	6	RAB32	rs2275606	146,918,950	Imputed	A	G	0.170	0.159	1.09 (0.85-1.40)	0.76
	5	IL12B	rs6863015	158,843,235	Imputed				QC failed		
	16	LOC388210	rs77061563	11,452,542	Imputed	T	C	0.434	0.437	0.99 (0.82-1.19)	0.97
	11	CCDC88B	rs663743	64,107,735	Imputed				QC failed		
	10	Intergenic	rs58600253	64,507,904	Imputed	T	C	0.117	0.121	0.96 (0.72-1.28)	0.74
	19	SIGLEC5	rs10414149	52,127,744	Imputed				QC failed		
	5	CDH18	rs73058713	20,374,213	Imputed	A	C	0.181	0.202	0.88 (0.69-1.11)	0.36
	9	1-Dec	rs10817758	118,143,933	Imputed	T	C	0.383	0.357	1.12 (0.92-1.35)	0.27
	1	BATF3	rs2221593	212,873,431	Imputed	T	C	0.122	0.120	1.02 (0.77-1.36)	0.78
TB	11	WT1	rs2057178	32,364,187	Genotyped	A	G	0.045	0.043	1.05 (0.67-1.65)	0.83
	8	ASAP1	rs4733781	131,296,767	Imputed	A	C	0.401	0.389	1.05 (0.87-1.27)	0.52
	18	Intergenic	rs4331426	20,190,795	Genotyped	G	A	0.013	0.022	0.58 (0.28-1.21)	0.14
	14	SMOC1	rs17175227	70,502,050	Genotyped	A	G	0.092	0.098	0.92 (0.67-1.27)	0.63
NTM	2	SLC11A1	rs2279014	219,261,176	Genotyped	T	C	0.236	0.239	0.98 (0.79-1.23)	0.89
	2	SLC11A1	rs17235409	219,259,732	Genotyped	A	G	0.088	0.107	0.81 (0.59-1.11)	0.19
	6	MICA	rs4418214	31,391,401	Genotyped	C	T	0.008	0.017	0.50 (0.21-1.19)	0.11
	6	MICA	rs2596542	31,366,595	Genotyped	T	C	0.307	0.355	0.81 (0.66-0.98)	0.033
	1	IL10	rs1800893	206,947,167	Genotyped	T	C	0.054	0.049	1.10	0.67

									(0.72-1.67)	
19	IL28B	rs8099917	39,743,165	Genotyped	G	T	0.094	0.104	0.89 (0.65-1.21)	0.45
6	TNFA	rs1799964	31,542,308	Genotyped	C	T	0.155	0.158	0.97 (0.75-1.26)	0.84
7	CFTR	rs213950	117,199,533	Imputed	A	G	0.357	0.359	0.99 (0.81-1.20)	0.91
7	CFTR	rs113857788	117,304,834	Imputed	C	G	0.045	0.028	1.62 (0.96-2.71)	0.11

SNP: single nucleotide polymorphism, TB: tuberculosis, NTM: nontuberculous mycobacteria, MAF: minor allele frequency, OR: odds ratio, CI: confidence interval, QC: quality control

Supplementary Table S3. Association of candidate SNPs at the replication stage

RSID	Chr	Position (hg19)	Nearest gene (Location / GENCODE)	Replication stage (591 MAC vs 718 Controls)						Combined (1066 MAC vs 1135 Controls)	
				MAF (Cases)	MAF (Controls)	OR (95% CI)	P	Call rate (%)	Bonferroni-adjusted P value at replication stage	OR (95% CI)	P
rs109592	16	23,768,710	CHP2 (Intron)	0.112	0.199	0.51 (0.41-0.63)	1.55E-09	1.000	9.30E-09	0.54 (0.45–0.63)	1.64E-13
rs11646605	16	23,793,439	CHP2 (23kbp 3')	0.157	0.241	0.59 (0.48-0.72)	1.39E-07	0.999	8.34E-07	0.58 (0.50–0.67)	3.62E-13
rs7200798	16	23,853,860	PRKCB (Intron)	0.293	0.345	0.79 (0.67-0.93)	4.77E-03	0.999	2.86E-02	0.77 (0.68–0.87)	5.74E-05
rs12140922	1	35,114,647	MIR552 (21kb 3')	0.063	0.061	1.05 (0.76-1.45)	0.762	1.000	1	0.75 (0.59-0.96)	2.28E-02
rs9668455	12	126,702,012	RP4-765H13.1	0.206	0.211	0.97 (0.80-1.17)	0.761	0.999	1	0.79 (0.68-0.91)	1.20E-03
rs12894463	14	98,999,346	RP11-1082A3.1 (18kb 3')	0.229	0.218	1.07 (0.89-1.28)	0.489	1.000	1	1.27 (1.10-1.46)	1.01E-03

SNP: single nucleotide polymorphism, MAC: *Mycobacterium avium* complex, MAF: minor allele frequency, OR: odds ratio, CI: confidence interval

Supplementary Table S4. Associations with rs109592 in Asian (Thai) TB

RSID	686 TB vs 771 Controls			
	MAF (Cases)	MAF (Controls)	OR (95% CI)	P
rs109592	0.227	0.212	1.09 (0.91–1.29)	0.361

TB: tuberculosis, MAF: minor allele frequency, OR: odds ratio, CI: confidence interval

Supplementary Table S5. SNPs associated with pulmonary NTM diseases

RSID	Chr	Position (hg19)	Minor Allele	Major Allele	MAF (Non-MAC)	MAF (MAC)	MAF (Controls)	P (MAC)	P (NTM)	OR in NTM GWAS (95%CI)
rs11646605	16	23,793,439	T	G	0.167	0.165	0.264	3.01E-07	1.81E-07	0.55 (0.44-0.69)
rs9668455	12	126,702,012	C	T	0.111	0.169	0.261	2.31E-06	4.30E-07	0.56 (0.45-0.70)
rs118088961	19	12,071,794	C	T	0.014	0.012	0.048	4.02E-06	2.44E-06	0.24 (0.12-0.45)
rs76665723	15	87,289,422	A	G	0.000	0.011	0.044	8.79E-06	2.48E-06	0.21 (0.11-0.43)
rs194797	16	23,780,839	C	T	0.111	0.129	0.209	6.06E-06	2.64E-06	0.55 (0.43-0.71)
rs700280	7	146,192,758	C	T	0.221	0.314	0.411	1.91E-05	3.57E-06	0.64 (0.53-0.77)
rs9669248	12	126,701,792	G	A	0.167	0.218	0.308	1.63E-05	4.60E-06	0.61 (0.50-0.76)
rs947389	10	108,254,348	A	C	0.389	0.328	0.236	1.74E-05	5.27E-06	1.61 (1.31-1.98)
rs2064474	6	33,073,463	A	G	0.208	0.222	0.139	6.13E-06	5.76E-06	1.76 (1.37-2.25)
rs700277	7	146,189,136	G	T	0.250	0.319	0.415	2.86E-05	7.29E-06	0.65 (0.53-0.78)
rs929259	17	32,465,809	C	T	0.417	0.378	0.484	6.24E-06	7.55E-06	0.65 (0.54-0.79)
rs2064478	6	33,072,266	T	C	0.208	0.220	0.138	8.37E-06	7.61E-06	1.75 (1.37-2.24)
rs78605884	6	138,928,169	C	T	0.028	0.055	0.014	4.68E-06	8.43E-06	3.83 (2.03-7.20)
rs910320	6	33,075,443	T	C	0.208	0.217	0.137	9.71E-06	8.49E-06	1.75 (1.36-2.24)
rs3097671	6	33,047,612	C	G	0.208	0.220	0.139	1.09E-05	9.95E-06	1.73 (1.36-2.22)

SNP: single nucleotide polymorphism, NTM: nontuberculous mycobacteria, MAF: minor allele frequency, OR: odds ratio, CI: confidence interval